2015

Ultracrepidarianism in Forensic Science: The Hair Evidence Debacle

David H. Kaye
Penn State Law

Ultracrepidarianism in Forensic Science: The Hair Evidence Debacle

David H. Kaye*

Abstract

For over 130 years, scientific sleuths have inspected hairs under microscopes. Late in 2012, the FBI, the Innocence Project, and the National Association of Criminal Defense Lawyers joined forces to review thousands of microscopic hair comparisons performed by FBI examiners over several of those decades. The results have been astounding. Based on the first few hundred cases in which hairs were said to match, it appears that examiners exceeded the limits of science in over 90% of their reports or testimony. The disclosure of this statistic has led to charges that the FBI faked an entire field of forensic science, placed pseudoscience in the witness box, and palmed off virtually worthless and scientifically indefensible evidence as scientific truth.

This essay disputes these interpretations of the 90+% figure. Based on some of the scientific literature on hair comparisons, the public descriptions of the hair review project that have emerged, and some of the confessions of scientific error that the FBI has issued, it reaches three conclusions: (1) associating two hairs by their physical features can be slightly probative of whether they originated from the same source; (2) the hair review project does not bear on the validity of these associations or the quality of the examinations; rather, it is supposed to flag cases in which examiners have overstated the power of a match to identify the source of the trace evidence; (3) some questionable determinations

* Associate Dean for Research and Distinguished Professor, School of Law, and Faculty Member, Graduate Program in Forensic Science, The Pennsylvania State University. I am grateful to Lynn Garcia and David Moran for information on the FBI's Microscopic Hair Comparison Analysis Review and to Max Houck and Barry Scheck for criticism of a draft of this paper.

have been issued, and the 90+% figure may not be a valid and reliable measure of overclaiming.

To promote a more complete understanding of the nature and extent of overclaiming, the review process should be made more transparent, and the materials it produces should be readily available for researchers and the public to study. New state or local evidence reviews should be designed with these concerns in mind. Finally, in all areas of forensic science, clear standards for presenting identification evidence without overclaiming should be devised, and training and monitoring programs should be implemented to ensure that laboratory personnel and prosecutors adhere to them.

Table of Contents

I. Introduction

Forensic-science practitioners commonly present findings regarding traces left at crime scenes, on victims or suspects, or on or in their possessions. Such trace evidence can take many forms. Physical traces such as fingerprints, striations on bullets, shoe and tire prints, and handwritten documents are common examples. Biological materials, such as blood, semen, saliva, and hairs also are fodder for the crime laboratory. Comparisons of a questioned and known sample can supply valuable information on whether a specific suspect is associated in some manner with a

crime.[1] Viewers of the acronymious police procedurals—*NCIS*, *CSI*, and *Law and Order SVU*—know all of this.

For decades, however, legal and other academics have questioned the hoary courtroom claims of absolutely certain identification of one and only one possible source of trace evidence.[2] In the aftermath of the 2009 report of a National Research Council committee,[3] these views have slowly gained traction in the forensic-science community. Indeed, in the popular press and among investigative reporters, the pendulum may be swinging in favor of uncritical rejection of once unquestioned forensic sciences. Recent months have seen an episode from *Frontline* presenting DNA evidence as "anything but proven";[4] they have included unfounded reports that as many as 15% of men and women imprisoned with the help of DNA evidence at trial are wrongfully convicted;[5] and award-winning journalists

---

1. Association is not necessarily causation. There may be innocent explanations for an impression or other trace that points to an individual. Thus, forensic scientists distinguish between "source level," "activity level," and "offense level" propositions. *See* COLIN G. G. AITKEN & FRANCO TARONI, STATISTICS AND THE EVALUATION OF EVIDENCE FOR FORENSIC SCIENTISTS 215–17 (2d ed. 2004) (discussing the relationship between association and causation); *cf.* DAVID H. KAYE ET AL., THE NEW WIGMORE: A TREATISE ON EVIDENCE: EXPERT EVIDENCE 601 (2d ed. 2011) (categorizing "hypotheses that might explain a reported match between samples said to be from the defendant and from a crime scene").

2. *See, e.g.*, David H. Kaye, *Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies*, 75 BROOK. L. REV. 1163, 1164 n.10 (2010) (citing some of the critical literature); D. Michael Risinger & Michael J. Saks, *A House with No Foundation*, 20 ISSUES SCI. & TECH. 35 (2003) (arguing that many forensic techniques used in courtroom proceedings rest on weak science); D. Michael Risinger et al., *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise"*, 137 U. PA. L. REV. 731 (1989) (examining the importance and shortcomings of expert handwriting identification evidence).

3. NAT'L RESEARCH COUNCIL, COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 47 (2009).

4. Some inaccuracies in the documentary are noted in David H. Kaye, *Frontline's Expose of DNA Testing: Yes and No*, FORENSIC SCI., STAT.& L. (June 24, 2015), http://for-sci-law-now.blogspot.com/2015/06/frontlines-expose-of-dna-testing-yes.html (last visited Sept. 24, 2015) (on file with the Washington and Lee Law Review).

5. *See* David H. Kaye, *DNA Evidence Causes Wrongful Convictions 15% of the Time!?*, FORENSIC SCI., STAT. & L. (July 3, 2015), http://for-sci-law-

have spread the word that the FBI "faked an entire field of forensic science,"[6] placed "pseudoscience in the witness box,"[7] and palmed off "virtually worthless" evidence as scientific truth.[8]

The last set of reports stem from an ongoing review of well over 20,000 cases in which the FBI issued reports on hair associations.[9] The review spans decades of hair comparisons, and it is showing so many questionable statements that the expert evidence on hair associations stands out as "one of the country's largest forensic scandals."[10] Its preliminary findings provoked prominent Senators to speak of an "appalling and chilling . . . indictment of our criminal justice system"[11] and to

---

now.blogspot.com/2015/07/dna-evidence-causes-wrongful.html (last visited Sept. 24, 2015) (debunking this assertion of the Rand Corporation) (on file with the Washington and Lee Law Review).

6. Dahlia Lithwick, *Pseudoscience in the Witness Box: The FBI Faked an Entire Field of Forensic Science*, SLATE (Apr. 22, 2015, 5:09 PM), http://www.slate.com/articles/news_and_politics/jurisprudence/2015/04/fbi_s_flawed_forensics_expert_testimony_hair_analysis_bite_marks_fingerprints.html (last visited Sept. 22, 2015) (on file with the Washington and Lee Law Review).

7. *Id*. These "shameful, horrifying errors" comprised "a story so horrifying . . . that it would stop your breath." *Id*.

8. Erin Blakemore, *FBI Admits Pseudoscientific Hair Analysis Used in Hundreds of Cases: Nearly 3,000 Cases Included Testimony About Hair Matches, A Technique That Has Been Debunked*, SMARTNEWS (Apr. 22, 2015), http://www.smithsonianmag.com/smart-news/FBI-Admits-Pseudoscientific-Hair-Analysis-Used-in-Hundreds-of-Cases-180955070/?no-ist (last visited Sept. 22, 2015) (quoting Ed Pilkington, *Thirty Years In Jail For A Single Hair: The FBI's 'Mass Disaster' of False Conviction*, GUARDIAN (Apr. 21, 2015), http://www.theguardian.com/us-news/2015/apr/21/fbi-jail-hair-mass-disaster-false-conviction) (on file with the Washington and Lee Law Review).

9. Norman L. Reimer, Exec. Dir., Nat'l Ass'n of Criminal Def. Lawyers, Microscopic Hair Comparison Evidence Audit of 2013, Address at Texas Criminal Justice Integrity Unit and Texas Forensic Science Commission Meeting (July 11–12, 2013) ("The FBI has discovered over 21,000 cases so far in which FBI hair examiners issued a report. Of the first 11,000 cases reviewed, reports finding a positive association between a questioned hair and a suspect's hair were discovered in approximately 1,650 cases.").

10. Spencer S. Hsu, *FBI Admits Flaws in Hair Analysis over Decades*, WASH. POST (Apr. 18, 2015), https://www.washingtonpost.com/local/crime/fbi-overstated-forensic-hair-matches-in-nearly-all-criminal-trials-for-decades/2015/04/18/39c8d8c6-e515-11e4-b510-962fcfabc310_story.html (last visited Sept. 22, 2015) (on file with the Washington and Lee Law Review).

11. *Id*. (quoting Sen. Richard Blumenthal (D-Conn.)).

call for a "root cause analysis"[12] of ubiquitous errors. A distressed Department of Justice and FBI joined with the Innocence Project and the National Association of Defense Lawyers not only to publicize these failings but also to call on states "to conduct their own independent reviews where . . . examiners were trained by the FBI."[13] Projecting the outcome of cases that have yet to be reviewed, post-conviction petitions refer ominously to "[t]housands of . . . cases the Justice Department now recognizes were infected by false expert hair analysis"[14] and "pseudoscientific nonsense."[15]

This Essay uses the hair scandal to differentiate two related problems with many types of forensic-science testimony. The first is the problem of foundation—what reasons are there to believe that hair or other analysts possess sufficient expertise to produce relevant evidence of associations between known and unknown samples? For years, commentators and some defense counsel have posed legitimate questions—with little impact in the courts—about the reliability and validity of physical comparisons by examiners asked to judge whether known and unknown samples are similar in enough respects—and not too dissimilar in other respects—to support a claim that they could have originated from the same individual.[16] To paraphrase Gertrude

---

12. *Id*. (quoting S. Judiciary Comm. Chairman Charles E. Grassley (R-Iowa) and Patrick J. Leahy (D-Vt.)).

13. Press Release, Fed. Bureau of Investigation, FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review (Apr. 20, 2015), http://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review.

14. Petition for a Writ of Certiorari at 2–3, Ferguson v. Steele, 134 S. Ct. 1581 (2014).

15. *Id*. at 19. Justice Breyer views the "errors" referred to in the press release as emblematic of "flawed forensic testimony" generally and a reason to hold the death penalty unconstitutional. Glossip v. Gross, 135 S. Ct. 2726 (2015) (Breyer, J., dissenting).

16. *See, e.g.*, Simon A. Cole, *Fingerprinting: The First Junk Science?*, 28 OKLA. CITY U. L. REV. 73, 76 (2003) (arguing that fingerprint examination is not forensic science but rather "experience-based practice"); Robert Epstein, *Fingerprints Meet* Daubert*: The Myth of Fingerprint "Science" Is Revealed*, 75 S. CAL. L. REV. 605, 607 (2002) (arguing that "a 'science of fingerprints' [is] an unfounded creation of law enforcement fingerprint examiners"); Clive A. Stafford Smith & Patrick D. Goodman, *Forensic Hair Comparison Analysis:*

Stein, is there enough "there there" to warrant any form of testimony about a positive association?[17] This is the existential question of whether "[t]he subject of an expert's testimony [is] 'scientific . . . knowledge.'"[18] Or, at the other extreme, is the entire enterprise ersatz—a "fake science" and a "worthless" endeavor? Part I of this Essay discusses this foundational question for physical hair comparisons. It rejects the harsh view that this endeavor is pure pseudoscience, like astrology, graphology, homeopathy, or metoposcopy, and argues that hair analysts generally possess *some* expertise in making associations.[19]

The second problem lies in the presentation of perceived associations. Even if there is a there there, are forensic-science practitioners staying within the boundaries of their demonstrated expertise? Or are they purporting to know more than they do? Part II of Essay shows that the FBI's revelations are limited to this issue of overclaiming. What the FBI has uncovered are expert assertions in one case after another that are said to outstrip the core of demonstrated knowledge. Such overclaiming is one form of scientifically invalid testimony,[20] but it is not the

---

Nineteenth Century Science or Twentieth Century Snake Oil?, 27 COLUM. HUM. RTS. L. REV. 227, 231 (1996) ("If the purveyors of this dubious science [forensic hair analysis] cannot do a better job of validating hair analysis than they have done so far, forensic hair comparison analysis should be excluded altogether from criminal trials."); *see* sources cited *supra* note 2 (discussing flaws in claims that forensic-science comparisons are entirely definitive).

17. "What was the use of my having come from Oakland . . . there is no there there." GERTRUDE STEIN, EVERYBODY'S AUTOBIOGRAPHY 289 (1937).

18. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 589–90 (1993).

19. It does not attempt to resolve the legal question of whether this foundational evidence is enough to justify admission of findings of associations at trial under the rules governing expert evidence. For a short argument that hair evidence is inadmissible under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, see 4 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 30:48, at 78–79 (2014). The argument faces a veritable wall of contrary case law. *See* Meskimen v. Commonwealth, 435 S.W.3d 526, 535–36 (Ky. 2013) (holding that the trial court did not abuse its discretion in admitting hair sample evidence); Commonwealth v. Chmiel, 30 A.3d 1111, 1141 (Pa. 2011) (claiming that hair microscopy is an accepted and reliable scientific technique); PAUL C. GIANNELLI & EDWARD J. IMWINKELRIED, SCIENTIFIC EVIDENCE § 24.03, at 651–57 (4th ed. 2007) (reporting that most courts are receptive to hair analysis).

20. For this vocabulary, see generally Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 VA.

equivalent of an entire invalid science. In addition, Part II questions the uncritical use of the proportion of problematic cases identified by the FBI as an estimate of the prevalence of scientifically indefensible testimony.

II. Matching Versus Guessing: Do Hair Examiners Have Anything to Offer?

Contrary to the more breathless accounts, the testimony of hair examiners—including the mistaken judgments that have contributed to wrongful convictions—did not emanate from a worthless pseudoscience. As a means of making positive associations between samples, they have been largely superseded by (generally) more definitive and (always) less subjective mitochondrial DNA typing.[21] Yet, mitotyping also proved that physical comparisons can be probative of identity. It is worth examining a few numbers to see how. In an intriguing study (usually cited as proof of the failings of microscopic hair comparisons),[22] FBI researchers looked at human hairs submitted to the FBI laboratory for analysis between 1996 and 2000 in which examiners had first conducted microscopic comparisons and tried to mitotype them afterward.[23] They counted the number of

---

L. REV. 1 (2009). *Cf.* Eric S. Lander, *Fix the Flaws in Forensic Science*, N.Y. TIMES, Apr. 21, 2015 ("The F.B.I. stunned the legal community on Monday with its acknowledgment that testimony by its forensic scientists about hair identification was scientifically indefensible in nearly every one of more than 250 cases reviewed.").

21. *See* DAVID H. KAYE, THE DOUBLE HELIX AND THE LAW OF EVIDENCE 226–38 (2010). Mitotyping has not eliminated the need for physical comparisons. *Id.* Hair comparisons are still used to screen cases for mitotyping, and in some instances, hairs with similar physical features lack sufficient DNA for mitotyping. *E.g.*, State v. Dotson, 450 S.W.3d 1, 36 (Tenn. 2014) (discussing DNA testing and its resulting evidence). In addition, mitotyping generally cannot distinguish between individuals who are maternally related.

22. *E.g.*, Reimer, *supra* note 9; RICHARD SAFERSTEIN, FORENSIC SCIENCE: AN INTRODUCTION 418 (2d ed. 2011).

23. Max M. Houck & Bruce Budowle, *Correlation of Microscopic and Mitochondrial DNA Hair Comparisons*, 47 J. FORENSIC SCI. 1 (2002).

concordant and discordant results of the two procedures to obtain the figures shown in Table 1.[24]

	Mitotyping +	Mitotyping −
Microscopy +	69	9
Microscopy −	0	17

Table 1. Comparison of Inclusions (+) and Exclusions (−) in Microscopic Hair Analyses with Those in Mitochondrial DNA Typing.

Every time mitotyping indicated a positive association, so did the previous microscopic inspections.[25] Thus, with mitotyping as the "gold standard,"[26] the sensitivity of the microscopic analysis in the cases studied was 69/(69+0) = 100%. Of course, some of these concordances might have been lucky guesses or influenced by extraneous knowledge of other facts about the case, but if these cases are thought of as a random sample from a much larger population of hairs collected in criminal investigations, one could say that microscopic hair examiners confronted with truly matching hairs from this population will declare them to match at least 95% of the time.[27]

For the 9 + 17 = 26 exclusions by mitotyping, the microanalysts reached the same conclusion as the DNA analysts—an exclusion—in seventeen cases, for a specificity of 65% (with a confidence interval of 44% to 83%). To put it another way, the physical hair analysis produced nine false positive associations out of the twenty-six cases in which mitotyping

24. Table 1 is derived from *id.* at 3, Table 2. It omits all cases in which hair examiners or DNA analysts could not reach a conclusion as to association.

25. In another fifteen cases, the mitochondrial match was preceded by an inconclusive microscopic comparison. Mitotyping in an additional thirteen cases produced a match when no microscopic comparison had been conducted, presumably because the hair sample was too limited. *Id.*

26. In clinical medicine, a "gold standard" is the best single test (or a combination of tests) for diagnosing a disease. A validation study for another method measures performance against this the gold-standard test. *E.g.*, Christopher D. Saudek et al., *A New Look at Screening and Diagnosing Diabetes Mellitus*, 93 J. CLINICAL ENDOCRINOLOGY & METABOLISM 2447 (2008).

27. The approximate 95% confidence intervals for the quantities given in the text were computed with MedCalc, https://www.medcalc.org/calc/diagnostic_test.php.

excluded the hairs and microscopic examinations had yielded either positive or negative associations (35%). However, there also were twenty cases of mitochondrial exclusions preceded by inconclusive microscopy. These were not false positive errors. Thus, it could be argued that the proportion of false positives was "only" 9/46 = 20%.[28] Whichever denominator applies, the nine false positives were cases in which microanalysis gave not only the wrong answer but an answer that could have contributed to a false conviction in the absence of confirmatory mitotyping.[29]

This false positive rate is cause for concern—and a factor that cuts against admitting findings of an association at trial.[30] Still, in the bulk of the cases of same-source hairs, the procedure gave the correct answer. It is not as if the analysts were flipping coins, tossing oracle bones, or dealing Tarot cards. Those methods would be just as likely to produce a declared association when hairs really are associated as when they are not. In contrast, the probability of an FBI examiner finding of a positive association when comparing hairs from the same individual exceeded the probability of this finding when comparing hairs from different individuals by a factor of 100/35 = 2.9 (with a confidence interval of 1.7 to 4.9). Evidence like this is weakly diagnostic of an association.[31]

---

28. In another thirty-four cases, no microscopic comparisons were undertaken. Houck & Budowle, *supra* note 23, at 3, Table 2.

29. For still other statistics on the correlation between the microscopic and mitochondrial tests, see MAX HOUCK & JAY SIEGEL, FUNDAMENTALS OF FORENSIC SCIENCE 310 (3d ed. 2015).

30. *See* Daubert v. Merrell Dow Pharms., 509 U.S. 579, 594 (1993) ("[T]he court ordinarily should consider the known or potential rate of error."); Paul C. Giannelli, *FBI Review of Microscopic Hair*, 29 CRIM. JUST. 31, 31 (reporting on the FBI's use of microscopic hair analysis).

31. By way of comparison, the positive likelihood ratio for the nitrite dipstick test and the Uriscreen test for urinary tract infection are .27/.06 = 4.5, and 1/.32 = 3.1, respectively. Anthony K. Akobeng, *Understanding Diagnostic Tests 1: Sensitivity, Specificity and Predictive Values*, 96 ACTA PÆDIATRICA 338, 339 (2006). The ratio for the PSA test for prostate cancer is approximately .7/.1 = 7. Carvell T. Nguyen & Michael W. Kattan, *Prediction Models in Prostate Cancer Diagnosis*, *in* PROSTATE CANCER DIAGNOSIS: PSA, BIOPSY AND BEYOND 85, 86 (J. Stephen Jones ed., 2012). Such values are disappointing for a clinical test, and no clinician would want to rely on these tests alone. Nevertheless, the numbers mean that the tests have some diagnostic value. Courts have been impressed with even smaller relative risks in epidemiologic studies. *See, e.g.*, *In*

A further indication that microscopic hair features have some value in separating same-source from different-source hairs comes from research published and debated in reputable journals since 1974. In that year, a Canadian forensic scientist[32] and a professor of mathematical statistics[33] described "An Attempt at Determining Probabilities in Human Scalp Hair Comparison."[34] In essence, hair analysts coded many features of a cross-section of about 9 hairs from each of 100 scalps.[35] Out of the recorded

---

re Actos (Pioglitazone) Prods. Liab. Litig., No. 6:12–cv–00064–RFD–PJH, 2014 WL 4364832 (W.D. La. Sept. 2, 2014) (treating a relative risk of less than three as helping establish a clinically significant hazard of bladder cancer from defendant's drug).

Even the sharpest critics of hair comparisons seem to regard them as useful for excluding suspects. *E.g.*, Editorial, *Junk Science at the F.B.I.,* N.Y. TIMES, Apr. 27, 2015, at A18 (discussing uses of hair comparison evidence). A method that reliably excludes even a fraction of possible suspects who are not the source increases the probability that an included suspect is the source. Thus, before more discerning DNA testing was available, courts recognized that even common blood types are admissible as evidence of identity. KAYE, *supra* note 21. If a hair has been shed from a person with type O blood (as established by a serologic test of the hair sheath), the positive likelihood ratio is about 2 (*see* LAURA DEAN, BLOOD GROUPS AND RED CELL ANTIGENS ch. 5 (2005), http://www.ncbi.nlm.nih.gov/books/NBK2267/), which is comparable to that for a microscopic association. On the other hand, blood grouping results are unlikely to be influenced by knowledge of extraneous facts about the case. Hair comparisons could be influenced by examiner expectations from facts about guilt or innocence. If the examiners had access to this information, some of their success in reaching the same outcomes as the mitotyping could be a reflection of that knowledge, rather than their professed skill at matching just the features in the hairs.

32. Barry Douglas Gaudette (1947–2003) "served 33 years as a Forensic Scientist with the Royal Canadian Mounted Police, most recently as Manager of the Canadian Police Research Centre, a collaboration of the National Research Council, Royal Canadian Mounted Police and the Canadian Association of Chiefs of Police." *"DNA" 2003 Obituary*, ONTARIO GENEALOGICAL SOC'Y PROVINCIAL INDEX (Oct. 4, 2003), https://www.ogs.on.ca/ogspi/2003/o2003dna.htm (last visited Aug. 30, 2015) (on file with the Washington and Lee Law Review).

33. Professor E. S. Keeping was head of the Department of Mathematics at the University of Alberta from 1954 to 1961. WALTER HUGH JOHNS, A HISTORY OF THE UNIVERSITY OF ALBERTA, 1908–1969 317 (1981).

34. Barry D. Gaudette & E. S. Keeping, *An Attempt at Determining Probabilities in Human Scalp Hair Comparison*, 19 J. FORENSIC SCI. 599 (1974).

35. *See id.* at 599 (stating that the actual number varied from six to eleven hairs selected to be representative of the variation seen in "80 to 100 hairs randomly selected from various regions of the scalp").

feature sets of all the 366,630 pairs of these nearly 900 hairs from different individuals, only 9 matched.[36] The probability of a false association was therefore 9/366,630 = 1/40,737. If p is the probability that a single hair from individual A will match a randomly chosen hair from individual B, then the probability of a non-match is $1 - p$. Hence, the probability of all 9 false inclusions is $(1 - p)^9 = 1/40{,}737$. Solving for p gives an approximate average probability of a false inclusion per hair of 1/4500.[37]

This particular number should not be regarded as an error rate for actual forensic hair comparisons. As the literature discussing and extending this study reveals, applying this average probability to casework requires heroic assumptions.[38] The hairs in the study may not have been representative of a population appropriate for casework, and the method used for matching those hairs differs from that in practice.[39] For a witness to claim this level of accuracy in a given case or to assure a court that hair analysts are generally this accurate would amount to "misinformation."[40] Nevertheless, the "average probability"[41]

---

36. *See id.* at 602 (comparing 861 hairs, giving rise to 370,230 distinct pairs in total and to 370,230 − 3,600 = 366,630 pairs from different individuals).

37. *See* THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS 65 (Stephen E. Fienberg ed., 1988) (citing Barry D. Gaudette, *A Supplementary Discussion of Probabilities and Human Hair Comparison*, 27 J. FORENSIC SCI. 279 (1982)) (stating that false matches in multiple comparisons are assumed to be independent, but "independence is not an important assumption since the use of the Bonferroni inequality yields an upper bound for the probability in question of 1/4,526").

38. *Id.* at 66; *see* Walter F. Rowe, *The Current Status of Microscopical Hair Comparisons*, 1 SCI. WORLD 868, 872–73 (2001) (reviewing the Gaudette-and-Keeping study and the subsequent literature); Smith & Goodman, *supra* note 16, at 246 (presenting concerns and modifying certain assumptions to obtain an average false-match probability of 1/261).

39. *But see* Houck & Budowle, *supra* note 23, at 2 ("[A] point-by-point, side-by-side comparison of the microscopic characteristics in the known sample with those in the questioned hair [used in the FBI study] . . . is more reliable than a simple subjective determination of the properties of an object [used in the Gaudette-Keeping study]."). The FBI study also involved a second examiner's verification, which should improve accuracy.

40. Fienberg, *supra* note 37, at 67; *see also* Colin G.G. Aitken & David A. Stoney, *Introduction*, *in* THE USE OF STATISTICS IN FORENSIC SCIENCE 23, 24 (Colin G.G. Aitken & David A. Stoney eds., 1991) ("Average probabilities . . . have a place in the assessment of the worth of a general area of evidence, such as head hairs, where frequency data are unavailable. However, it is very

analysis helps demonstrate that the physical features of hair provide information on true associations.[42]

Consequently, summarily to condemn the entire field as "junk science" from "so-called experts" that "would not make it within shouting distance of a peer-reviewed journal"[43] is ill-informed. That being said, there are significant legal arguments for excluding evidence of hair associations. Although microscopes are fine scientific instruments, and much is known about the biology of hair, the process of inferring an association is far from standardized. Hair examiners have always been clear that "[t]he conclusions drawn from a forensic hair comparison are subjective and result from an outgrowth of experience and judgment. Outside of personal observations acquired through the accumulated experiences of thousands of hair examinations, little data exists [sic] to aid the examiner in assessing the significance of a hair comparison."[44]

---

misleading to quote average probabilities in the courts as a measure of the value of the evidence.").

41. AITKEN & TARONI, *supra* note 1, at 133.

42. *Id.* at 229–31; *see* Colin G.G. Aitken, *Populations and Samples*, *in* Aitken & Stoney, *supra* note 40, at 51, 72 ("[T]he research of Gaudette & Keeping is valuable, and it showed that hair comparison can distinguish hairs from different sources."); *cf.* A.J. Jeffreys et al., *Individual-Specific 'Fingerprints' of Human DNA*, 316 NATURE 76 (1985) (using a similar calculation to make extreme claims of individuality for multilocus DNA profiles). A few additional experiments into the accuracy of microscopic hair examinations, as well as an extension of the Gaudette-Keeping study, are summarized in Barry D. Gaudette, *Evidential Value of Hair Examination*, *in* FORENSIC EXAMINATION OF HAIR 243, 250–53 (James Robertson ed., 1999).

43. Editorial, *supra* note 31, at A18.

44. *See* Richard E. Bisbing, *The Forensic Identification and Association of Human Hair*, *in* FORENSIC SCIENCE HANDBOOK 184, 210 (Richard Saferstein ed., 1982)

> The merits of each comparison must be independently assessed before the strength of the conclusion is expressed. Factors such as the number of unknown hairs found to be similar to the controls, the number of microscopic features observed, the presence of unusual characteristics, the condition of the specimens, and the number and completeness of the exemplars must all be taken into account;

see also Edward J. Imwinkelried, *Forensic Hair Analysis: The Case Against Underemployment of Scientific Evidence*, 39 WASH. & LEE L. REV. 41, 62 (1982) (acknowledging the subjectivity of scientists when categorizing hair samples).

But "experiences" are not scientific observations,[45] and true expertise does not develop without consistent feedback on performance.[46] As far as scientific research into hair comparisons goes, the average-probability studies indicate that trained examiners can code physical features that tend to differentiate hairs from different individuals. The FBI data on the correlation between mitotyping and microscopy indicate that, notwithstanding the absence of articulated standards for measurement and interpretation, analysts can make correct classifications at levels exceeding pure chance. However, the FBI data come from one small, retrospective study and the examiners may have had the benefit of extraneous information indicating whether the hairs would be expected to match or not. Are the publications enough to establish a reasonably accurate, generally applicable, and low enough error rate to warrant admission of the evidence? Are discrepancies between microscopy and mitotyping randomly distributed, or are there variables that make the perceived positive and negative morphological associations—the inclusions and exclusions—more accurate or less accurate in identifiable situations? Considering that the FBI has been combining microscopy with mitotyping for years, it should not be hard to replicate and extend the 2002 study to address some of these questions. Thus, the better argument against admissibility of all microscopic association evidence is not that hair morphology provides no meaningful information, but that more needs to be done to demonstrate and quantify the validity of the judgments of

---

45. *See* Michael Baum, *The Controlled Trial and the Advance of Reliable Knowledge*, 287 BRIT. MED. J. 1216, 1216 (1983) ("The trouble with experience as a way of approximating to reliable knowledge is that all of us tend to reinterpret each individual experience in the light of a previously held conceptual framework.").

46. *See* KAYE ET AL., *supra* note 1, § 10.3.3

> Numerous studies have found that without quick and accurate feedback on correct and incorrect judgments, experience does not produce expertise and experts routinely overestimate their skills . . . . Casework in forensic handwriting analysis, latent fingerprint identification, toolmark identification, and other patterns and impression evidence comparisons rarely involve . . . feedback based on ground truth. The argument that the judgments of these analysts are valid merely because the practitioners have had specialized training or ample experience therefore is unimpressive.

associations before accepting the risk that judges or juries will read too much into hair comparisons no matter how carefully they are presented.[47]

Courts have almost always rejected this argument in favor of the view that hair evidence is not so falsely impressive that it should be kept from factfinders. No matter the outcome of this evidentiary question,[48] the ongoing FBI review does not probe the scientific validity of the discipline; rather, it finds a pervasive problem with the overstated manner in which (weakly) probative evidence has been presented. But how clear is it that examiners are exceeding the limits of science nine times out ten?

III. Association Versus Definitive Identification: How Do Hair Examiners Testify?

There is reason to think that the FBI review project is casting a wide net in classifying reports and testimony as exceeding scientifically established limits of our knowledge about positive associations. Erring in the direction of being overly inclusive may be a reasonable strategy for notifying defense counsel of possible concerns, but it is sure to overestimate the prevalence of error in cases with hair evidence. Reviewers are not asking whether, on balance, reports and testimony are limited to a report of a positive association. Rather, they seem to be searching for *any* statements that, read in isolation and left unqualified, would *suggest* that an individual is the source of the hair in question.

To indicate the resulting ambiguity in the statistic on scientifically erroneous testimony, this Part describes the broad criteria articulated by the FBI and its partners and then draws on Department of Justice letters to prosecutors and defense

47. *See* Giannelli, *supra* note 30, at 32 ("[T]he evidence may have little probative value, while at the same time be quite misleading."). The risk that jury will overvalue scientific evidence of limited accuracy is the raison d'être of the special scrutiny it should receive under cases such as *Frye v. United States* and *Daubert v. Merrell Dow Pharmaceuticals*. *See generally* KAYE ET AL., *supra* note 1 (discussing topics and issues surrounding admitting scientific evidence).

48. *See supra* note 19 and accompanying text (explaining this legal question in greater detail).

attorneys and on other sources for examples of laboratory statements that have been deemed erroneous. It concludes that the high rate of reported errors includes an unknown number of contestable judgments. Unless the underlying reports and transcripts are made available for more rigorous research, an independent, scientifically credible evaluation of any statistics on the prevalence of errors and their significance is not feasible.

A. The Hair Comparison Analysis Review Project

In July 2012, the Department of Justice (DOJ) announced that it and the FBI had begun to identify "historical cases for review where a microscopic hair examination conducted by the FBI was among the evidence in a case that resulted in a conviction."[49] But "details of how the new FBI review will be conducted [were] unclear."[50]

Over the next year, the Department and the Bureau "worked closely" with the Innocence Project (IP), the National Association of Criminal Defense Lawyers (NACDL), and pro bono lawyers "in determining the scope, protocols and implementation of [a] review [of] more than 2,000 cases that were processed by the FBI [before] 2000."[51] The most complete published explication of the protocol for determining whether statements "were scientifically invalid"[52] or exceeded "the permissible limits of the science of hair

---

49. Spencer S. Hsu, *Justice Dept., FBI to Review Use of Forensic Evidence in Thousands of Cases*, WASH. POST (July 10, 2012), http://www.washingtonpost.com/local/crime/justice-dept-fbi-to-review-use-of-forensic-evidence-in-thousands-of-cases/2012/07/10/gJQAT6DlbW_story.html (last visited Sept. 3, 2015) (quoting Department of Justice spokeswoman Nanda Chitre) (on file with the Washington and Lee Law Review).

50. *Id.*

51. *Innocence Project and NACDL Announce Historic Partnership with the FBI and Department of Justice on Microscopic Hair Analysis Cases*, INNOCENCE PROJECT (July 18, 2013, 12:00 AM), http://www.innocenceproject.org/news-events-exonerations/press-releases/innocence-project-and-nacdl-announce-historic-partnership-with-the-fbi-and-department-of-justice-on-microscopic-hair-analysis-cases (last visited Sept. 3, 2015) [hereinafter *Innocence Project*] (on file with the Washington and Lee Law Review).

52. *Id.*

microscopy"[53] comes from the NACDL.[54] This article explains that the reviewers look for "three types of error."[55] A "Type I" error occurs when an "examiner stated or implied that the evidentiary hair could be associated with a specific individual *to the exclusion of all others*."[56] How the reviewer decides what is implied is entirely undefined.[57] A Type II error exists when an

> examiner assigned to the positive association a *statistical weight or probability* or provided a likelihood that the questioned hair originated from a particular source, or an opinion as to the likelihood or rareness of the positive association that *could lead the jury to believe* that valid statistical weight can be assigned to a microscopic hair association.[58]

---

53. Norman L. Reimer, *The Hair Microscopy Review Project: An Historic Breakthrough For Law Enforcement and a Daunting Challenge for the Defense Bar*, CHAMPION (July 2013), http://www.nacdl.org/Champion.aspx?id=29488 (last visited Sept. 3, 2015) (on file with the Washington and Lee Law Review).

54. *See id.* (explaining that after reviewing the case materials, the FBI will send its determination to the IP and NACDL for their review, with the aim of notifying defense counsel or the defendant of erroneous statements).

55. *Id.*

56. *Id.* (emphasis added).

57. Illustrating a Type I error, the article offers an easy hypothetical example:

> I found brown, Caucasian head hairs on two items of clothing, the sports coat, and a pair of slacks that were reported to me as belonging to [the defendant]. Now, these hairs matched in every observable microscopic characteristic to that known hair sample of DEC [the decedent] and consistent with having originated from her. In my opinion, based on my experience in the laboratory and having done 16,000 hair examinations, my opinion is that those hairs came from DEC.

Id.

58. *Id.* (emphasis added). The accompanying example is as follows:

> Q: Now, based on your training and experience and your expertise in the field, and based on your knowledge of hair transfer and hair comparison, and based on the work done in this case, do you have an opinion, within the degree of scientific certainty, as to whether or not the pubic hair found in the underpants of [victim] came from [defendant]?
>
> A: I would say that it would be a very high degree of probability that it does. Or to reverse it, I would say the chances of it being from somebody else, other than Mr. XX, would be highly unlikely at best.

Finally, a Type III error arises when an "examiner cites the *number* of cases or hair analyses worked in the lab and the number of samples from different individuals that could not be distinguished from one another as a predictive value *to bolster* the conclusion that a hair belongs to a specific individual."[59]

Soon after entering into this unprecedented agreement, the FBI halted its review, allegedly because "[n]early every criminal case reviewed . . . included flawed forensic testimony."[60] The FBI attributed the hiatus to "a vigorous debate that occurred within the FBI and DOJ about the appropriate scientific standards we should apply when reviewing FBI lab examiner testimony—many years after the fact."[61] Apparently, the FBI lost the debate. The Deputy Attorney General ordered the Bureau to proceed "under the original terms."[62]

That was late in July of 2013. On April 20, 2015, the FBI issued a stunning press release acknowledging that "examiners' testimony in at least 90 percent of [268] trial transcripts the Bureau analyzed as part of its Microscopic Hair Comparison Analysis Review contained erroneous statements."[63] The

---

59. *Id.* (emphasis added). The example is: "Now over the last 12 years, I personally have looked at hairs from about 10,000 different people, and over that time, I've only had two occasions out of the 10,000 people where I had hairs from two different people that I could not separate them." *Id.*

60. *See* Spencer S. Hsu, *Federal Review Stalled After Finding Forensic Errors by FBI Lab Unit Spanned Two Decades*, WASH. POST (July 29, 2014), http://www.washingtonpost.com/local/crime/federal-review-stalled-after-finding-forensic-errors-by-fbi-lab-unit-spanned-two-decades/2014/07/29/04ede880-11ee-11e4-9285-4243a40ddc97_story.html (last visited Sept. 2, 2015) (reviewing some 160 cases) (on file with the Washington and Lee Law Review).

61. *See id.* ("'Working closely with DOJ, we have resolved those issues and are moving forward with the transcript review for the remaining cases,' the FBI said.").

62. *Id.*

63. FBI, *supra* note 13. The press release explains:

> The government identified nearly 3,000 cases in which FBI examiners may have submitted reports or testified in trials using microscopic hair analysis. As of March 2015, the FBI had reviewed approximately 500 cases. The majority of these cases were trials and the transcript of examiner testimony was reviewed. Some of these cases ended in guilty pleas, limiting the review to the original lab report. In the 268 cases where examiners provided testimony used to inculpate a defendant at trial, erroneous statements were made in

disclosure ignited the media firestorm about "fairytale hair evidence,"[64] a "discredited tool,"[65] and an "elite FBI forensic unit"[66] that was "wrong in 96 pct. of cases."[67]

Yet, it is not so easy to know what to make of the 90+ percent figure. Clearly, it does not mean that FBI analysts were wrong to report and testify to—in some manner—a positive association in the 268 cases of testimony for the prosecution.[68] Although there undoubtedly are cases in which FBI analysts announced that questioned and known hairs could have come from the same source when, in fact, the hairs came from different sources,[69] determining "ground truth" in the thousands of cases is all but impossible. In any event, this task is not part of the review. The issue before the reviewers is not the scientific validity of hair

---

257 (ninety-six percent) of the cases.

See also Hsu, *supra* note 10 (noting that a less-detailed report of these results came from the DOJ, FBI, IP, and NACDL the day before).

64. Sean Allocca, *In the Public Eye: FBI's Fairytale Hair Evidence and the Real-Life Consequences*, FORENSIC MAG. (July 7, 2015, 10:03 AM), http://www.forensicmag.com/videos/2015/07/public-eye-fbis-fairytale-hair-evidence-and-real-life-consequences?et_cid=4663527&et_rid=454869145&location=top (last visited Sept. 4, 2015) (on file with the Washington and Lee Law Review).

65. Joseph A. Slobodzian, *Hair Analysis a Discredited Tool: FBI Review Shows It Was Wrong in 96 Pct. of Cases*, PHILADELPHIA INQUIRER, May 12, 2015, at A1.

66. Hsu, *supra* note 10.

67. Slobodzian, *supra* note 65.

68. *See FBI/DOJ Microscopic Hair Comparison Analysis Review*, FBI, http://www.fbi.gov/about-us/lab/scientific-analysis/fbi-doj-microscopic-hair-comparison-analysis-review (last visited July 5, 2015) (outlining that the FBI continues to refer to visual and microscopic comparisons as perfectly capable of yielding "probative associations" and emphasizes that "microscopic hair comparison analysis is a valid scientific technique still conducted by the FBI Laboratory") (on file with the Washington and Lee Law Review).

69. *See* Spencer S. Hsu, *Kirk Odom, Who Served 20 Years for 1981 D.C. Rape, Is Innocent, Prosecutors Say*, WASH. POST (July 10, 2012), http://www.washingtonpost.com/local/crime/2012/07/10/gJQAUjZNbW_story.html (last visited Sept. 4, 2015) (acknowledging scientific errors after DNA evidence proved another man was the culprit) (on file with the Washington and Lee Law Review); David H. Kaye, *The FBI's Worst Hair Days*, FORENSIC SCI., STAT. & L. (July 31, 2014), http://for-sci-law-now.blogspot.com/2014/07/the-fbis-worst-hair-days.html (last visited Sept. 4, 2015) (discussing the prosecution's overstated presentations of microscopic hair associations in two cases) (on file with the Washington and Lee Law Review).

comparison as such—although, as noted in Part I, that is indeed open to question. The issue is not whether the specific judgments in the cases were gross departures from the norm for ascertaining when samples could have originated from a common source. It is not whether claims of a positive association are, *ipso facto*, scientifically unacceptable and therefore should not have been made.

To the contrary, the NACDL and the IP

> agreed with the FBI that an examiner's testimony concerning the relationship between two hairs is appropriate [to show] that a contributor of a known sample could be included in a pool of people of unknown size, as a possible source of the hair evidence (without in any way giving probabilities, as an opinion to the likelihood or rareness of the positive association, or the size of the class) . . . .[70]

The IP further explained

> [I]t is possible to conduct hair microscopy and find similarities among various samples. But it appears that in many cases the FBI analysts were overstating the significance of these similarities, often leaving juries with the false impression that a hair recovered from the crime scene must have come from the defendant and could not have come from anyone else.[71]

In sum, the only issue in the review is whether the examiners were claiming or implying substantially more than a "possible source" association. "Possible source" testimony is the kind that hair examiners have always said they could legitimately provide. The professional literature has long acknowledged that microscopic characteristics are not useful for "absolute personal identification", but only for "determining whether or not a questioned hair *could have* originated from a particular individual."[72] Experts have testified accordingly.[73]

---

70. Reimer, *supra* note 53.

71. *Innocence Project*, *supra* note 51 (quoting Peter Neufeld, co-founder and co-director). Of course, this concession may have been a strategic one, rather than a full-throated endorsement of microscopic comparisons.

72. *See* Douglas W. Deedrick & Sandra L. Koch, *Microscopy of Hair Part 1: A Practical Guide and Manual for Human Hairs*, FORENSIC SCI. COMM. (Jan. 2004), https://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/jan2004/research/2004_01_research01b.htm (last visited

Given the cautions in the professional literature, what is shocking is that so many examiners so often went beyond the limits of the knowledge that their discipline claimed.

But how often did this occur? The FBI Hair Comparison Review does not seek to provide a precise and scientifically defensible estimate of the prevalence of ultracrepidarian

---

Sept. 4, 2014) ("The variability and distribution of the microscopic characteristics are useful in determining whether or not a questioned hair *could have* originated from a particular individual [but] hair comparisons do not constitute a basis for absolute personal identification.") (emphasis added) (on file with the Washington and Lee Law Review); JOHN W. HICKS, MICROSCOPY OF HAIR: A PRACTICAL GUIDE AND MANUAL 7 (1977)

> [T]he examiner may conclude (1) that the hairs are consistent or similar and could have come from the same source, (2) that the hairs are dissimilar and did not come from the same source or (3) that the hairs possess characteristics which are not sufficiently defined to arrive at a meaningful conclusion.;

id. at 4 ("[T]he possibility cannot be dismissed that there may be two hair samples whose ranges of variation overlap and, therefore, a positive identification cannot be made."); *id.* at 41 ("[H]airs do not possess a sufficient number of unique individual microscopic characteristics to be positively identified as having originated from a particular person to the exclusion of all others."); Houck & Budowle, *supra* note 23, at 2 ("Microscopic comparison of hairs has never been considered a positive form of identification."); NAT. RES. COUNCIL, *supra* note 3, at 156 ("[A] conclusion of a 'match' means only that the hair could have come from any person whose hair exhibited—within some levels of measurement uncertainties—the same microscopic characteristics, but it cannot uniquely identify one person."); *Forensic Human Hair Examination Guidelines*, FORENSIC SCI. COMM. (Apr. 2005), http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/april2005/index.htm/standards/2005_04_standards02.htm (last visited Sept. 3, 2015) ("Individualization is the process of attempting to determine whether a given hair came from one particular (person) source to the exclusion of all other sources. This is not possible with forensic microscopical hair comparison.") (on file with the Washington and Lee Law Review). *But see* Bisbing, *supra* note 44, at 203 (allowing for the possibility of "individuality . . . only under extremely unusual circumstances").

73. *See, e.g.*, State v. Cameron, 674 P.2d 651, 655 (Wash. 1983) ("[T]he expert stated only that the hair 'could have originated from this individual.'"). *But see* Tani Watkins et al., *The Science of Forensic Hair Comparisons and the Admissibility of Hair Comparison Evidence:* Frye *and* Daubert *Considered*, MCCRONE GROUP (Mar. 2, 2004), https://www.mccrone.com/mm/the-science-of-forensic-hair-comparisons-and-the-admissibility-of-hair-comparison-evidence-frye-and-daubert-considered/ (last visited Sept. 4, 2015) ("Most experts testify that the likelihood of finding someone else with indistinguishable hair is remote, a rare event.") (on file with the Washington and Lee Law Review).

testimony. The purpose of the exercise is to notify defendants whose conviction might be tainted by overstated testimony, along with the prosecutors, so that they can pursue any appropriate relief. That is a laudable purpose, but it does not demand scientific precision. It is better to be overinclusive than underinclusive in screening for cases of possible injustice. The Hair Comparison Analysis Review is a massive effort to spot all cases with statements that should not have been made. As we soon shall see, the reviewers are not evaluating the nature of the testimony as a whole by considering whether potentially misleading statements were corrected or explained appropriately, and the reviewers could be using contestable criteria for classifying individual statements as erroneous.

B. How Are the Reviewers Classifying Statements?

A White House task force to protect students from sexual assault began its report with a statistic designed to underscore the seriousness of the problem: "One in five women is sexually assaulted in college."[74] But other estimates are "all over the map . . . ,"[75] in part because of variations in the definitions of sexual assault and the wording of the specific survey questions that operationalize these definitions.[76] To convey the campus meaning of "sexual assault," some colleges supply explicit examples to cover "the full range of nonconsensual sexual activity."[77]

---

74. THE WHITE HOUSE TASK FORCE TO PROTECT STUDENTS FROM SEXUAL ASSAULT, NOT ALONE: THE FIRST REPORT OF THE WHITE HOUSE TASK FORCE TO PROTECT STUDENTS FROM SEXUAL ASSAULT 2 (2014).

75. Joanne Lipman, *The Toughest Issue on (Any) Campus*, YALE ALUMNI MAG., July–Aug. 2015, at 41; *see also* Sofi Sinozich & Lynn Langton, Bureau of Justice Statistics, Rape and Sexual Assault Victimization Among College-Age Females, 1995–2013, Dec. 2014, at 2 (detailing the National Crime Victimization Study's findings that the victimization rate was an order of magnitude less (4.7 per 1,000 for females ages 18 to 24 enrolled in post-secondary schools)).

76. Sinozich & Langton, *supra* note 75, at 41.

77. *Id.*

Just as one cannot make sense of statistics on the prevalence of campus sexual assault without knowing the details of how the estimates were obtained, one cannot conclude that 90% or more is a sound estimate of overclaiming without knowing the specifics about statements that the FBI classifies as Type I, II, and III errors. The generic definitions are hardly self-applying, and the hypothetical examples from the NACDL do not span the kinds of testimony that actually occur.

To appreciate the problem, it may be helpful to consider the reports or testimony in a few cases in which the FBI has confessed error. Ambiguities in the definitions would be academic if all cases were clear, but the life of the law is more complicated. The outcome for some cases is obvious enough; for instance, in one Florida case, the examiner testified that "[i]t's highly unlikely that the pubic hair originated from anybody else than Mr. Grady, and it's highly unlikely that the head hair originated from anybody else . . . the chances of that would be almost nonexistent."[78] This is a clear "Error Type I"—an express or implied assertion "that the evidentiary hair could be associated with a specific individual to the exclusion of all others."

A slightly less obvious example comes from a wrongful conviction case in the District of Columbia that helped precipitate the FBI's review project, where an analyst "said . . . the hair in the stocking came from Tribble."[79] Such testimony is a source attribution, but it is not clear from this description whether the analyst identified Tribble "to the exclusion of all others." This phrase was extremely popular among analysts of impression and patterns (like fingerprints and tool marks) who believed that their discipline studies characteristics that can exist in their particulars in only one object in the universe. Of course, the words "to the exclusion" are logically redundant.[80] If the analyst

---

78. Reimer, *supra* note 9 (displaying transcript of Michael Malone's direct testimony).

79. Maurice Possley, *Santae Tribble*, NAT'L REGISTRY OF EXONERATIONS (July 21, 2014), http://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3926 (last visited Sept. 11, 2015) (on file with the Washington and Lee Law Review).

80. *See generally* Simon Cole, *Individualization Is Dead, Long Live*

believed that "the hair . . . came from Tribble," then he must have believed that it did not come from anyone else. However, one reasonably can believe that a named individual is the source of a trace—because that is the most likely conclusion—without believing it is impossible for anyone else to have been the source (which is most likely what "to the exclusion" was supposed to mean).[81] Thus, there is an ambiguity in the meaning of "Error Type I." How explicit and firm must the analyst be in excluding all other individuals as possible contributors of the hair?[82]

Now consider *United States v. MacDonald*,[83] one of the most storied cases in the annals of crime.[84] There, an examiner testified that "this hair—in conducting a comparison examination with the comparison microscope—microscopically matched the head hairs of Colette MacDonald."[85] This statement was not deemed erroneous, since it made no assertion beyond the bare fact that the questioned and known hairs matched. But upon being asked on cross-examination why he was not concerned about the fact that the questioned hair was entangled with a

---

Individualization! Reforms of Reporting Practices for Fingerprint Analysis in the United States, 13 L., PROBABILITY & RISK 117 (2014).

81. *See* Kaye, *supra* note 5 (discussing the meaning of individualization and uniqueness in forensic science).

82. One case indicates that "to the exclusion of all others" is not a limitation at all. In one of two double-murder charges for which Willie Manning was convicted, an FBI agent testified that normally he could make a source attribution "to some degree of certainty." Manning v. State, 726 So.2d 1152, 1180 (Miss. 1998). In this case, he did not do so; rather, he said that because the hairs recovered from a car "were only fragments," he could make no comparison and "was limited to a determination as to the racial characteristics of the hair." Manning v. Epps, 695 F. Supp. 2d 323, 380 (N.D. Miss. 2009). The Department of Justice issued a confession-of-scientific-error letter anyway. David Kaye, *Justice Breyer in* Glossip v. Gross *on "Flawed Testimony from an FBI Hair Examiner,"* FORENSIC SCI., STAT. & L. (July 16, 2015), http://for-sci-law-now.blogspot.com/2015/07/justice-breyer-in-glossip-v-gross-on.html (last visited Sept. 11, 2015) (on file with the Washington and Lee Law Review).

83. Nos. 3:75–CR–00026–F, 5:06–CV–00024–F, 2015 U.S. Dist. LEXIS 64572 (E.D.N.C. May 18, 2015).

84. *See generally* David H. Kaye, *No Relief for Jeffrey MacDonald After FBI Declares It "Exceeded the Limits of Science" with Hair Analysis*, FORENSIC SCI., STAT. & L. (May 23, 2015), http://for-sci-law-now.blogspot.com/2015/05/no-relief-for-jeffrey-macdonald-after.html (last visited Sept. 11, 2015) (on file with the Washington and Lee Law Review).

85. *Id.* (quoting testimony of Paul Stombaugh).

thread, the examiner explained that he was examining it only to reach a "conclusion" about "its origin."[86] How it had been stored was irrelevant to that question. The FBI, the NCDL, and the IP all classified the exchange as a Type I error—an express or implied assertion "that the evidentiary hair could be associated with a specific individual to the exclusion of all others." Never mind that the examiner had not testified that the hair originated from Colette MacDonald or anyone else. In this instance, the reviewers' judgment that this examiner made a source attribution of any kind seems questionable.

Moving to some examples of Type II errors, an examiner in an Arkansas case testified "it is possible for two different people's hair to exhibit the same uniqueness, but it's very rare. . . . [I]t would be a very low degree of probability of it happening."[87] This statement does not seem to supply a *statistical* weight or probability, but it unequivocally expresses "an opinion as to the likelihood or rareness of the positive association." As such, it clearly exemplifies one category of a Type II error.[88] Compare this with the report in *MacDonald* stating that "[a] forcibly removed Caucasian head hair . . . exhibits the same microscopic characteristics as hairs in the K2 specimen. Accordingly, this hair is consistent with having originated from Kimberly MacDonald, the identified source of the K2 specimen."[89]

This, the FBI concluded, was also a Type II error. But why? Again, the examiner had not "assigned to the positive association a statistical weight or probability."[90] Additionally, he had not "provided a likelihood that the questioned hair originated from a

86. *MacDonald*, 2015 U.S. Dist. LEXIS 64572, at *6; *see also* Kaye, *supra* note 84 (detailing the cross-examination and testimony with further specificity).

87. *See* Reimer, *supra* note 9 (displaying an excerpt from the transcript in *Arkansas v. Spencer*).

88. *See also* Duckett v. State, 568 So.2d 891, 893 (Fla. 1990) ("While other experts could not reach a conclusion by comparing that hair with Duckett's pubic hair, Michael Malone, an FBI special agent [concluded] that there was a high degree of probability that the pubic hair found in her underpants was Duckett's pubic hair.").

89. United States v. MacDonald, Nos. 3:75–CR–00026–F, 5:06–CV–00024–F, 2015 U.S. Dist. LEXIS 64572, at *14 (E.D.N.C. May 18, 2015).

90. Reimer, *supra* note 53.

particular source . . . ,"[91] nor had he expressed an "opinion as to the likelihood or rareness of the positive association."[92] Apparently, examiners can say that two hairs share the same features—they "match," or they are "consistent with" one another—but must not add the obvious (and scientifically undeniable) fact that this observation (if correct) means that the hairs could have had the same origin or that they are "consistent with" this possibility.[93] And even if they add that "[h]air comparisons are not a basis for personal identification," as this examiner did, they have, in the FBI's view, "exceeded the limits of science."[94]

Of course, one can criticize phrases like "consistent with" and "match" as creating an unacceptable risk that—in the absence of clarification on direct examination, cross-examination, or by judicial instruction—jurors will think the words connote a source attribution. But this would make their use a Type I error, which does not seem to be the FBI's position.[95] More importantly, arguments of this sort stray from determinations that an examiner has made statements that "exceed the limits of science" to judgments that an examiner has made statements that are scientifically acceptable but prone to being misunderstood.

---

91. *MacDonald*, 2015 U.S. Dist. LEXIS 64572, at *5.

92. *But cf. supra* note 57 and accompanying text (contrasting this example with a traditional Type II error).

93. Although a statement that two hairs with matching features are "consistent with" the hypothesis of a common origin seems well within the bounds of science, the phrase—and various related ones—are far from ideal. *See* I. W. Evett et al., *The Impact of the Principles of Evidence Interpretation on the Structure and Content of Statements*, 40 SCI. & JUST. 233, 237 (2000) (explaining why the given phrase tends to "obscure" rather than clarify).

94. *Id.* (quoting the Department of Justice letter in the case).

95. One might argue that it is poor science to consider consistency with only the prosecution's same-source hypothesis and not the defense's different-source hypothesis. An expert who wants to use "consistent with" terminology might testify that (1) the match is consistent with both hypotheses, but (2) it lends more support to the former as compared to the latter, and (3) there is currently no adequate basis indicating how much more support it provides for the same-source hypothesis. This comparative statement about the weight of the evidence seems scientifically justified. If it were not, the relevance of any testimony of matching features would be in doubt. 1 MCCORMICK ON EVIDENCE § 195, at 996–97 (Kenneth S. Broun ed., 7th ed. 2013).

To be sure, this latter danger is important to the law. As noted in Part I, it should inform rulings of admissibility under Federal Rules of Evidence 403 and 702. It is a reason to regulate the manner in which experts testify to scientifically acceptable findings, as some courts have done.[96] Laboratories themselves should adopt and enforce policies to ensure that reports and testimony avoid terminology that is known to convey the wrong impression. But it is misleading to include scientifically acceptable, but psychologically dangerous, phrasing in the counts of scientifically erroneous statements. Case-review projects ought to flag all instances in which examiners have not presented their findings as they should have, but reports ought to differentiate between statements that directly "exceed the limits of science" and those that risk being misconstrued in a way that would make them "exceed the limits of science." One size does not fit all.

IV. Conclusion

For over 130 years, scientific sleuths have been inspecting hairs under microscopes to ascertain whether particular hairs could have come from the same person.[97] This branch of forensic science can supply evidence that is weakly probative of identity.[98] The Microscopic Hair Comparison Analysis Review initiated in 2012 does not alter this conclusion. Instead, the Review presupposes that there is a scientific basis for some testimony but that FBI examiners have been known to exceed what this foundation supports. Despite the legitimate challenges that can be raised to the rather standardless nature of microscopic hair associations, claims that the case review project has shown them to be a mere pseudoscience that produces wrong answers over 90% of the time rest on a misunderstanding of the nature of the review.

---

96. *See generally* KAYE ET AL., *supra* note 1 (discussing the admission of scientific evidence)

97. *See* Imwinkelried, *supra* note 44, at 41 ("This year [1982] marks the one hundredth anniversary of the use of scientific hair evidence in American Prosecutions.").

98. *See generally supra* Part I.

Of course, as a screening test for overclaiming, the review project need not meet the publication standards that social science would demand for a valid and reliable application of the taxonomy of errors that the FBI and its partners created. Undertaking that kind of study would not be trivial. It would require a detailed and relatively unambiguous protocol to guide the reviewers in their determinations.[99] It would include some demonstration that the reviewers were applying the guidelines reliably. And it would entail creating a dataset of reports and transcripts for inspection by other researchers. Ironically, while the State Department is releasing thousands of pages of Hillary Clinton's emails,[100] the reports and transcripts held by the Justice Department are not readily available for independent scrutiny.

Nevertheless, the ongoing Hair Analysis Comparison Review already has delivered a crucial lesson to the forensic science community and the criminal justice system. It seems likely—and it is deplorable—that many hair examiners have overstated the probative value of their findings of positive associations—and not just in the rare and exceptional case. Although just how flagrant all the overstatements have been is unclear and no precise estimate of the prevalence of overclaiming is available, there is no reason to wait for an exact statistic or for more nuanced measures. Steps can and should be taken in all jurisdictions and all areas of forensic science to reduce the incidence of overclaiming. The Organization of Scientific Area Committees needs to develop clear standards for the limits on testimony about hair and other forms of scientific identification evidence.[101]

---

99. After this Essay was in press, the FBI responded affirmatively to a Freedom of Information Act request for written instructions received by its examiners. I plan to discuss this internal guidance document in later publications on the presentation of trace evidence.

100. *See* Polly Mosendz, *State Department Releases 7,000 Pages of Hillary Clinton's Emails*, NEWSWEEK (Aug. 31, 2015, 9:04 PM), http://www.newsweek.com/state-department-releases-7000-pages-hillary-clintons-email-367458 (last visited Sept. 22, 2015) (discussing the State Department's release of 7,000 of Clinton's emails) (on file with the Washington and Lee Law Review).

101. Nat'l Institute of Standards and Technology, ORG. OF SCI. AREA COMMS. (OSAC) (Feb. 7, 2014), http://www.nist.gov/forensics/osac.cfm (last visited Sept.

Laboratories need to implement or maintain training and continuing education programs designed to ensure that analysts: (1) understand what has been scientifically established in their field; (2) appreciate that impressions of accurate judgments on their part or on the part of their colleagues are not acceptable measures of validity or strength of evidence; and (3) know how to provide answers at trial that reflect these understandings even when pressed to make stronger claims. Testimony should be monitored so that expert witnesses receive feedback when they approach or cross a clear line established by the laboratory for testimony. Parallel training and monitoring programs for prosecutors should ensure that they (1) do not make arguments that overstate the scientifically established power of the evidence; and (2) do not frame questions that invite the witness do so, directly or by implication.[102] Finally, as the shock brought on by the 90%+ statistic triggers statewide hair testimony (or other) review programs, the different types of overclaiming can be defined more precisely, and appropriate examples from the federal review can be employed in new instructions to reviewers. In these ways, the hair evidence debacle can be transformed from a source of cynicism to a force for maturation.

22, 2015) ("NIST is working with the forensic science community to establish the new Organization for Scientific Area Committees (OSAC). OSAC will coordinate development of standards and guidelines for the forensic science community to improve quality and consistency of work in the forensic science community.") (on file with the Washington and Lee Law Review).

102. Judicial orders also can be framed to preclude overclaiming, and jury instructions can be issued to correct transgressions.